

Обучение машинного перевода без параллельных текстов*

Рябов Ф. А. Бахтеев¹ О. Ю. Стрижов² В. В.

¹Московский физико-технический институт

²Вычислительный центр им. А. А. Дородницына ФИЦ ИУ РАН

Данная задача посвящена исследованию алгоритма обучения машинного перевода без параллельных предложений. Использование параллельных текстов для задачи машинного перевода требует слишком большой базы предложений всех переводимых языков, что является ресурсоемкой задачей для некоторых пар непохожих языков. Особенностью исследуемого алгоритма является то, что для перевода используется кодирование и декодирование текста во внутреннем представлении. Данный алгоритм использует единую модель нейронной сети Seq2Seq для перевода с одного языка на другой и обратно. Цель данного исследования заключается в том, чтобы сделать вектора скрытых пространств этих двух моделей как можно более похожими. Для демонстрации работоспособности метода будет использован вычислительный эксперимент машинного перевода между двумя похожими языками: русским и украинским.

Ключевые слова: *машинный перевод, нейросеть, Seq2Seq.*

1 Введение

Решается задача оптимизации системы машинного перевода без использования параллельных предложений, так как для некоторых пар языков получение таких пар предложений, а также и само обучение является ресурсоемкой задачей [2].

Существует ряд подходов к построению систем машинного перевода [3, 4, 8]. Предлагается использовать рекуррентные нейронные сети с короткой и долгой памятью и нейронные сети, в которых реализовано механизм внимания (attention). В данном методе используются нейронные сети, которые осуществляют перевод в два этапа: кодирование и декодирование. Такой метода называется Seq2Seq [6, 7].

Данная работа посвящена последнему методу последовательного перевода. Предлагается с помощью первой рекуррентной нейронной сети, основанной на долгой памяти перевести входящую последовательность в вектор, а с помощью второй перевести этот вектор в выходную последовательность на нужном нам языке [1]. Предлагается модель, которая оптимизируется таким образом, чтобы скрытые пространства для векторов предложений двух языков совпадали. Данный метод позволяет обучить нейронную сеть переводу с одного языка на другой, в связи с использованием ей предыдущего опыта и наличию у нее памяти и внимания.

Эксперименты и анализ качества предложенного метода проводится на паре языков "русский-украинский" с помощью метрик BLEU (Bilingual evaluation understudy) для проверки качества систем машинного перевода [5].

2 Постановка задачи

Рассматриваемая модель отображает предложения из обоих языков в одно общее векторное пространство и из этого пространства в предложения первого или второго языка.

*Работа выполнена при финансовой поддержке РФФИ, проект №00-00-00000. Научный руководитель: Стрижов В. В. Авторы: А.В. Грабовой, О.Ю. Бахтеев, В.В. Стрижов, Eric Gaussier, координатор Малиновский Г.С. Консультант: Бахтеев О. Ю.

Она делает это с помощью кодировщика f и декодировщика g , в качестве которых выступают рекуррентные нейронные сети.

Обозначим через $\mathfrak{D}^{\text{src}} = [\mathbf{s}_1^{\text{src}}, \dots, \mathbf{s}_{m_{\text{src}}}^{\text{src}}]$ корпус предложений из первого языка, через $\mathfrak{D}^{\text{tgt}} = [\mathbf{s}_1^{\text{tgt}}, \dots, \mathbf{s}_{m_{\text{tgt}}}^{\text{tgt}}]$ – корпус предложений из второго языка, не являющиеся параллельными. Так как во время обучения не используются пары параллельных предложений, то требуется добиться того, чтобы автокодировщики вели себя похожим образом. Этого можно добиться, потребовав сходство скрытых векторных пространств. Для этого предлагается использовать сеть-дискриминатор q , которая по векторному представлению h некоторого предложения определяет, какому языку оно принадлежит. Функция потерь записывается таким образом, чтобы оптимизировать дискриминатор r для распознавания представлений предложений, а автокодировщик – для генерации максимально похожих представлений слов из разных языков. Для борьбы с переобучением кодировщика используется механизм зашумления. Пусть $\sigma(x)$ – результат наложения шума на слово x . Оптимизировать будем следующую функцию:

$$L_{AE} = ||d(e(\sigma(x))) - x||^2$$

Пусть дана какая-то модель слабого перевода \hat{g} . На втором шаге функция потерь будет иметь вид:

$$L_{TR} = ||d(e(\hat{g}(e(x)))) - x||^2$$

Пусть дана модель D , различающая скрытые представления векторов предложений из двух языков. На последнем шаге оптимизируем дискриминатор, чтобы он различал представления векторов разных языков в скрытом пространстве:

$$L_{ADV} = \log p(\text{lang} = \text{src} | \text{Encoder}(x)) + \log p(\text{lang} = \text{tgt} | \text{Encoder}(y))$$

В итоге нужно минимизировать следующую функцию:

$$L = a * L_{AE} + b * L_{TR} + c * L_{ADV} \longrightarrow \min$$

здесь a, b, c – калибруемые гиперпараметры.

Литература

- [1] Kyunghyun Cho, Bart Van Merriënboer, Dzmitry Bahdanau, and Yoshua Bengio. On the properties of neural machine translation: Encoder-decoder approaches. *arXiv preprint arXiv:1409.1259*, 2014.
- [2] Alexis Conneau, Guillaume Lample, Marc’Aurelio Ranzato, Ludovic Denoyer, and Hervé Jégou. Word translation without parallel data. *arXiv preprint arXiv:1710.04087*, 2017.
- [3] Philipp Koehn. *Statistical machine translation*. Cambridge University Press, 2009.
- [4] Philipp Koehn, Hieu Hoang, Alexandra Birch, Chris Callison-Burch, Marcello Federico, Nicola Bertoldi, Brooke Cowan, Wade Shen, Christine Moran, Richard Zens, et al. Moses: Open source toolkit for statistical machine translation. In *Proceedings of the 45th annual meeting of the ACL on interactive poster and demonstration sessions*, pages 177–180. Association for Computational Linguistics, 2007.
- [5] Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. Bleu: a method for automatic evaluation of machine translation. In *Proceedings of the 40th annual meeting on association for computational linguistics*, pages 311–318. Association for Computational Linguistics, 2002.

- [6] Ilya Sutskever, Oriol Vinyals, and Quoc V Le. Sequence to sequence learning with neural networks. In *Advances in neural information processing systems*, pages 3104–3112, 2014.
- [7] Ron J Weiss, Jan Chorowski, Navdeep Jaitly, Yonghui Wu, and Zhifeng Chen. Sequence-to-sequence models can directly translate foreign speech. *arXiv preprint arXiv:1703.08581*, 2017.
- [8] Yonghui Wu, Mike Schuster, Zhifeng Chen, Quoc V Le, Mohammad Norouzi, Wolfgang Macherey, Maxim Krikun, Yuan Cao, Qin Gao, Klaus Macherey, et al. Google’s neural machine translation system: Bridging the gap between human and machine translation. *arXiv preprint arXiv:1609.08144*, 2016.

Поступила в редакцию